DOCUMENT RESUME

ED 212 668 TM 820 087

AUTHOR Ellett, Frederick S., Jr.; And Others
TITLE Philosophical Remarks about Educational Evaluation: Evaluation and Values [and] The Foundations of Educational Evaluation. Studies in Evaluation and Decision Making. Work Unit 3.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Jun 78
NOTE 74p.; For related documents, see TM 820 088-090.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Educational Assessment; Educational Philosophy; *Evaluation; *Theories; *Values
IDENTIFIERS Empiricism; Evaluation Problems

ABSTRACT

A philosophical criticism of some of the major views of educational evaluation is presented. In particular, it is argued that the major views have serious limitations that result from their lack of understanding of questions of value, especially ethical value. Current views of educational evaluation need to draw upon the philosophical writings on values in order for their views to be justifiable. Evaluation as a practical activity is grounded on empirical facts and theories of value. Educational evaluation theories belong either to the descriptive (or empirical) realm or the normative realm, and there are various forms of normative educational evaluation activity, each having distinctive goals, purposes, principles, standards of excellence, obligations, and duties. An understanding of the logical structure of theories of educational evaluation will not only enhance discussion of the aforementioned, but will lead to greater defensibility in its theory-guided practice. (GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED212668

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

DELIVERABLE: June 1978

STUDIES IN EVALUATION AND DECISION MAKING

WORK UNIT 3

The Application of Philosophical Methods of Inquiry to the Practice of Educational Evaluation

Frederick S. Ellett, Jr.
Project Director

PHILOSOPHICAL REMARKS ABOUT EDUCATIONAL EVALUATION: EVALUATION AND VALUES

F. S. Ellett, Jr., G. D. Fenstermacher, B. Shumener

THE FOUNDATIONS OF EDUCATIONAL EVALUATION

Frederick S. Ellett, Jr.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G.S. Gray
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 087

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California - Los Angeles

June 1978

"The activity which is the subject of this report was supported in part or in whole by the National Institute of Education, Department of HEW. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred."

Philosophical Remarks about Educational Evaluation:

Evaluation and Values

F. S. Ellett, Jr., G. D. Fenstermacher, B. Shumaner

This essay presents a philosophical criticism of some of the major views of educational evaluation. In particular, we will argue that the major views have serious limitations that result from their lack of understanding of questions of value, especially ethical value. We will argue that current views of educational evaluation need to draw upon the philosophical writings on values in order for their views to be justifiable. We will argue that evaluation as a practical activity is grounded on empirical facts and theories of value.

Some Troubling Remarks

Alkin has facetiously stated that, surely, we are all aware of and "convinced of the value of evaluation," for he concludes his essay by stating that in certain situations "the context and decision factors are so predetermined that it can inevitably be said that no one needs it and no one cares" (Alkin, 1975, pp. 201, 211).

Now it is an important conceptual point that, in Western society at least, evaluation is a valued activity. It is an activity which people regard, rightly or wrongly, as being worthwhile. It is an activity which people look to for answers regarding the worth of the objects under study.

It is disturbing, therefore, to find that Alkin remarked elsewhere that "some evaluators readily accept the role of 'federal form filler-outers' and 'state statute satisfiers' and willingly do what is necessary to satisfy those requirements. That, too, is evaluation" (Alkin, 1976).

If it were proper to identify as evaluation activities which have dubious, or perhaps pernicious value, then an evaluation activity would not always be worthwhile. In some cases, at least, certain forms of evaluation would not be a good thing and could conceivably be a bad thing. This is especially true in those cases where the public awaits answers to questions of concern, yet receives completed federal forms instead. If Alkin is right, then one might properly conclude that certain forms of evaluation are not needed and that people are right in not caring about or in resisting such evaluations. Perhaps, then, people are wrong in thinking that evaluation is a good thing. Perhaps people ought to re-examine the limitations and the strengths of activities called evaluation.

Whether evaluation is a worthwhile activity depends in part on how evaluation is understood. Whether one should do evaluations is dependent in part on what the purpose of evaluation is. Is it helpful to say that mere federal form filling-out and state statute satisfying is evaluation? We turn now to address this and related questions.

The View of Stufflebeam and Guba

We shall begin by discussing the view which has come to be widely accepted by people writing about these things. (Many views of evaluation have been offered, and they do conflict with one another.)

In the monograph on evaluation produced by Phi Delta Kappa, Stufflebeam and Guba say that evaluation is the process of delineating, obtaining, and providing useful information for judging decision alternatives (Stufflebeam & Guba, 1971, p. xxv). The writers believed that this view of evaluation represents the essential or necessary features of evaluation, not just the accidental or contingent features.

Stufflebeam and Guba were not merely offering their personal view of evaluation, i.e., they were not merely saying, "This is what we regard as evaluation." They were doing much more. What Stufflebeam and Guba attempted to do was to provide their reader with an explication of evaluation. Theirs was an attempt to delineate the kinds of activities which must take place whenever one evaluates.

That the writers believed that they were doing this is made clear by looking at the passages which follow:

> The basis for this definition (view of evaluation) is to be found in dictionary definitions of its key terms. Among other definitions, evaluation is defined as the ascertainment of value, and decision, as the act of making up of one's mind. The need to make up one's mind connotes the existence of competing alternatives and to select one over the other(s) requires ascertaining their relative values. Hence, evaluation is the process of ascertaining the relative values of competing alternatives. Although less specific, this definition of evaluation is consistent with the one proposed in this book (Stufflebeam & Guba, p. xxv).

Clearly, then, the writers believed that they were providing the essential or necessary characteristics of evaluation. Without such features or characteristics, an activity could not be an evaluation activity.

There are good reasons for trying to provide an explication of the essential features of evaluation. It is surely desirable from a theoretical point of view to have a clear idea--as clear an idea as possible--about the nature of evaluation activities. Otherwise, one's empirical studies and practical recommendations are likely to go astray. The

second reason, which is related to the first, is that evaluation plays an important part in many activities of ordinary life. If one were merely to present his view of evaluation, there would be the ever-present danger that people would confuse the personal viewpoint with the more common or shared viewpoint. There are enough difficulties without adding the difficulties which might arise from such misunderstandings. It seems to us, then, that the writers were right in trying to provide, at least as a first step, the essential features of evaluation.

There are some problems, however, with the basis of the Stufflebeam-Guba view of evaluation. For the writers go on to state that the need to make up one's mind connotes the existence of competing alternatives and that to select one over the other(s) requires ascertaining their relative values. Thus, such a view implies that if one merely thinks he has alternatives before him he is not making up his mind--he is not making a decision. According to their view, a person can make a decision only if two (or more) alternatives actually exist. Now we think this point is well taken, but what about cases in which the alternatives exist but the person is not aware of them? Surely we would want to say that the person cannot make a decision if he is unaware of the alternatives. Therefore, Stufflebeam and Guba should add the necessary condition for deciding that the person must be aware of the alternatives. It is proper to say that a person has made a decision only if there are two (or more) alternatives actually open to him and only if he is aware of the two (or more) alternatives.

Second, and more importantly, they hold that deciding upon one alternative over the other(s) requires ascertaining their relative values. That is, one decides upon one of the alternatives only if he

ascertains the relative values. Their view has the consequence that if one were to make a mistake about the relative values of the alternatives, then one could not be deciding upon one of the alternatives at all. The view holds, then, that it is inconceivable for someone to decide to do something when he has made a mistake in his attempt to ascertain the relative values. Now people make good decisions and bad decisions, but they make decisions nonetheless. To make a decision is in part to deliberate about taking alternative courses of action.

The reason for the mistake is that the writers have wrongly equated evaluation with deliberation. When one makes up one's mind--more precisely, when one deliberates--one tries, before acting, to determine which is the best course open to him with a view to entering on it. One is "ready," in some sense, to act upon the course which he thinks, after deliberation, is the best one. Needless to say, people can deliberate about which course of action is best and still not discover the best course. When one deliberates, he tries to evaluate the alternatives open to him, but he can deliberate and yet fail to determine the relative values of the alternatives, fail to evaluate the alternatives. When one deliberates, he is looking for what are the best reasons for and against the courses open to the agent. Looking is one thing; finding what one is looking for is another (Baier, 1965, p. 42).

Stufflebeam and Guba were entitled to conclude that deliberation is necessarily the process whereby one tries to determine the relative values of the alternatives. In deliberation, one tries, before acting, to evaluate the alternatives to determine the best course for him. Thus, an evaluation of the alternatives is what one tries to achieve in the process of deliberating about the courses of action open to the

agent. It is important to notice two things about this statement. First, it is not a necessary or essential feature of evaluation; it is a necessary statement about the purpose of deliberation. One can do an evaluation without deciding what action to take. Second, unless the reader already knows full well what deliberation is, and hence he implicitly knows about evaluation, the statement will not help him understand evaluation. At best it will remind him of something he already knows. At worst, it will not increase his understanding at all. If someone were to ask you what a scientific theory is, it would probably not be helpful to tell him a scientific theory is what a scientist tries to achieve.

We conclude, therefore, that the basis offered in defense of the view of Stufflebeam and Guba is unsound. Rather than adequately explicating evaluation, it confuses evaluation with deliberation.

"Evaluation" and Supervenience

Upon completion of an evaluation, the person (or group) draws conclusions which in a rough way come to the following: the program is good (or bad), desirable (or undesirable), or worthwhile (or not worthwhile), and so on. When a person reaches such a conclusion, we are always entitled to ask, "What is good (or bad), say, about that?" This question must always admit of an answer. In philosophical jargon, "good" and related words refer to supervenient characteristics of situation. One may always legitimately be asked when one has called something a "good" something, "What is good about that?" Now to answer this question is to give the properties in virtue of which we call it good. Sometimes we cannot specify just what it is that makes it good, but there must always be something (Hare, 1952, p. 131).

Because "good," say, is a supervenient epithet, one may always legitimately be asked when one has called something a "good" something, "What is good about that?" Ultimately, the answer appeals to certain standards, rules, or principles. Such standards or principles are often called ethical (or value) principles, or simply values. To be entitled to apply the word "good," then, a person must have a reason for thinking that the object is good. But reasons are no respectors of persons. There is a thrust toward objective justifiability. We are entitled to ask whether the person's reason is a good reason for calling the object good. Ultimately, his reason must show that the program satisfies certain ethical principles--that it has certain values. Our point, then, is that there is a conceptual connection between evaluation and value.

Another Argument in Favor of the View of Stufflebeam and Guba

At this point, one may grant that the basis offered by Stufflebeam and Guba is unsound, but one may argue that the view can be supported in another way.

Recall that their explication states that evaluation is the process of delineating, obtaining, and providing useful information for judging decision alternatives. According to their explication, the process of evaluation is concerned with judging alternative courses of action. Someone may point out in favor of their view that it does make the essential connection between evaluation and values. Furthermore, someone may point out that Stufflebeam and Guba's explication correctly identifies those standards, rules, or principles which are appropriate. Here is how Stufflebeam and Guba put it:

If evaluation is a process that furnishes information useful in guiding decision making, the first operational step is the identification of the most useful information. Two things must be known:

1. What decision alternatives are to be considered (for it is about these alternatives that the information must be obtained).
2. What values or criteria will be applied (for the collected information must bear on these).

These two sets of specifications can be obtained by the evaluator only in interaction with his client (Stufflebeam & Guba, 1971, p. 42).

Such a passage leads one to believe that it is the client's values or criteria which should be used. But the writers are unsure about this. They also write "that he need not accept the client's definition, no matter how foolish, or limit himself to the client's criteria, no matter how irrelevant. His task is to criticize the client's formulation" (Stufflebeam & Guba, p. 42). They fail, however, to specify which standards, principles, or criteria are to be applied here. And later they write about the "valuing process" this way:

> Value to whom? Obviously, to the client or decision maker. To the extent that the decision maker reflects a community of interest within and outside the institution he represents, his values may be a composite of various interest or reference groups. The political basis for arriving at value relates to system maintenance or power, so that decision makers, responsible for defining a value base, inevitably use those persons with a power stake in the system to define value (Stufflebeam & Guba, p. 112).

Apparently they believe that the appropriate principle is that one ought to serve the power structure. But why is such a view defensible? What if the community is committed to irrational and immoral practices? Don't those who lack power in the system deserve some consideration in the evaluation?

Their view is not helped by their specifying that at least four kinds of expectations should be considered: (1) the expectations (institutional values) of the Board of Education, (2) expectations of parents, (3) expectations of decision maker's co-workers, and (4) the personal or private values of the decision maker (Stufflebeam & Guba, p. 113). The same problems arise here, too.

According to Stufflebeam and Guba, then, the appropriate standards and criteria--values--are to be determined by finding out what certain groups of people expect, like, or prefer. It is important to realize that this view is shared by others. Popham expressed a similar view in the following way:

> The most popular educational needs-assessment models emphasize the accumulation of a considerable amount of preference data, typically from a variety of different educational clientele, such as pupils, teachers, and parents. This preference data deals with which educational goals various groups would most like to see a given instructional system achieve (Popham, 1975, p. 65).

Although Popham seems to disagree with Stufflebeam about which groups should be consulted, he does agree with Stufflebeam that to find the appropriate values and criteria one has to determine what certain people most like (Popham, 1975, p. 69). Popham did add, later on, that the decision maker didn't have to rely solely on such preference data.

According to Popham, the decision maker should make a personal judgment as to the worth of the objectives under consideration, for he may be able to make "more enlightened judgments" as to the worth of various educational objectives (Popham, 1976, p. 69). As Stufflebeam and Guba had done, Popham, too, fails to specify what standards or criteria should be used to make a "more enlightened" judgment.

Worthen and Sanders do not believe that decision makers' preferences are always logical, rational, or intelligent revelations of value. According to their view of evaluation, the curriculum evaluator would first identify the curriculum goals, and using input from appropriate reference groups, determine whether or not the goals were good for the students, parents, and community served by the curriculum. But goodness is quickly equated with preference. And Worthen and Sanders wrote that "equating values with preference" had precedent in economics (Worthen and Sanders, 1973, pp. 21, 25, 26). Such a move by Worthen and Sanders shows a lack of understanding of economics and of the nature of values. For surely it does matter in evaluation whether a person's preferences are informed, rational, or moral.

Stake has written that the evaluator ought to go beyond the stated objectives of the curriculum in order to find out from the people involved whether <u>they</u> <u>see</u> the project's objectives as highly appropriate (Stake, 1973, p. 314). Although Stake's position presents an interesting variation of the theme described above, his position is open to the same criticisms.

In a recent book on curriculum evaluation, Klein and Tyler wrote about values in the following way:

> Values permeate all thinking, feeling, and acting. Whether we are selecting a novel for a friend, thinking about the functioning of an institution, or feeling unhappy about a friend's behavior, our values are an integral part of our mental and psychological processes (Tyler & Klein, 1976, p. 26).

All of these writers share the belief that the appropriate values for evaluation can simply be identified with what a person (or a certain group) most likes or most prefers. They hold that when we speak of a value or of values, we refer to what is valued, felt to have value, desired, or thought to be good. The expressions "the group's values" and "his value system" refer to what a group and what a man value or think to be good. Therefore, "value" means being valued (or preferred or desired), and "good" means being thought good or most desired.

We must make an important point about this use of the term "value" and its cognates. The term "value" can properly be used in such a way. The meaning of the term in one of its senses is such that to say that someone values something is to say that someone thinks highly of, esteems, or prizes that something. But we must remember that the term "value" is used in other ways. In particular, the term "value" is used to mean what has value or is valuable, or good, as opposed to what is regarded as good or valuable. Thus, "values" means "things that have value," or "things that are good." A value is that which is desirable or worthy of esteem. To equate the former meaning of the term with the latter meaning of the term is to commit a form of the naturalistic fallacy. One cannot derive a valuable from a value.

The distinction between the two meanings of the term, then, is an important one. It is one we recognize in our daily lives. We sometimes

value--regard as valuable--things we later come to see are not valuable. We sometimes make mistakes in our thinking or fail to think carefully about whether something is valuable, and upon reflection and comparison we come to see that we were wrong. Sometimes our views are uninformed; sometimes swayed by prejudice and bias. We are, or should be, careful to criticize our values to determine whether what we esteem is indeed worthy of esteem. This point may be summed up roughly by the slogan, "What interests the public may not be in the public interest."

In thinking about the nature of evaluation of actions, one should keep the following point well in mind.

> Every human action is in principle subject to practical and moral evaluation. It always makes sense to ask whether what human beings are doing (or going to do) is prudent or wise, satisfactory, intelligent, right or wrong. All action is directed, rightly or wrongly, correctly or mistakenly, towards ends held by the agent to be right. Accordingly, it always makes sense to ask whether these ends are properly evaluated or whether there is reason to suppose that their choice has been misguided or evil (Black, 1975).

It is always logically permissible, and sometimes ethically obligatory, to ask whether what the person values is really valuable. There is no logical barrier to going, and in some cases an obligation to go, beyond what the person thinks is valuable to ask whether it is valuable.

We conclude, therefore, that current writers on educational evaluation have committed a serious mistake by equating what people regard as valuable with what is valuable or worthy of esteem. Insofar as views similar to that of Stufflebeam and Guba imply that the appropriate value is what people regard as valuable, the view is mistaken and likely to be

harmful to practice. Evaluation is not finding what people regard the value of something to be, but finding what is its value.

An Argument Concerned with Direct Impact

Those writers, such as Marvin Alkin, who are worried about the direct or immediate impact of evaluations, often suggest that the first and major job of the evaluator is to play to the major contextual "needs" of the given situation (Alkin, 1975, 1976). Instead of just looking at the decision alternatives, the evaluator is urged to look at the real purpose of doing the evaluation, at those who commissioned the evaluation, and at the range of actual possibilities for conducting an evaluation. According to this view, the important point is that if the evaluation is to in any way be useful--if it is to make a direct and immediate impact upon decision making--the evaluator must be aware of what is the major contextual starting point and he must "play to" it. The evaluation is to be directed toward certain factors of the situation so that the evaluation will move the real decision maker, the agent in power.

Such suggestions are in some ways related to the Stufflebeam-Guba view which held that an evaluation, among other things, provides information which will be used by the decision maker. It follows from this view that if the information provided were not used, then the activity would be not an evaluation.

Clearly, there is something wrong with this view. We can provide somebody with good reasons, even conclusive reasons, for doing something, and yet the person can fail to act on them. If we criticize a governmental program because it brings about harmful results, our evaluation is still valid even if the government is unresponsive to the just criticism.

Perhaps Alkin is merely offering a view of evaluation which urges that the information provided be used. He does feel, however, that evaluators will be better prepared to provide the right kind of information if they understand who the real decision makers are, the decisions they make, and the constraints upon their decision making. The word "right" here, however, seems to mean little more than "useful in playing to the major contextual needs of the situation," to those in control (Alkin, 1975, p. 211).

A question which should immediately arise is whether an activity which "plays to the major contextual setting" is worth doing at all. That the evaluation "influenced" the decision process is, so far, not a good reason in favor of conducting such an activity. For what if the "evaluation" report were used by the existing powers to "legitimate" their denying educational opportunities to minority or disadvantaged students? The "evaluation" may well delineate, obtain, and provide information useful to those in power. Those in power may put the information to "good use" in the maintenance of their arbitrary privileges and authority. Surely such activities would not be justifiable or worthwhile. That policy which is justifiable must address more than the interests of those in authority and power, and it must consider those interests in a certain way.

Perhaps observations about the political realities led Campbell to the conclusion that evaluations should be conducted only where the political climate is propitious, where the political community is poised to move with the results (Campbell, 1975, p. 95). Apparently Campbell concluded that if the political community were not prepared to move with the justified conclusions of the evaluation, then the evaluation would

not be worth doing. This view is mistaken for it assumes that an evaluation is worth doing only if the political community is ready to use the evaluation findings. Campbell has overlooked the fact that evaluations can have indirect effects. The evaluation may produce a public criticism of the institution's practices that eventually will bring changes to the institution by social or legal means.

In certain cases information may be perceived as a good by those in power because of its usefulness for their purposes. Philosophers call such goods (or values) utility values. A utility value is a thing that is good because of its efficacy in reaching some end. But why should one stop an evaluation of a program when one has ascertained that the program has certain utility values or when one has ascertained that the valuation will not be immediately acted on by those in power? Social criticism, as we know it, would cease to exist under such conditions. Many evaluations have rather indirect effects on institutional programs and structure. In situations where human needs are being neglected or where humans are being harmed, one is entitled and, indeed, one is morally obligated to evaluate those purposes which give rise to the utility values. Whether utility values are good or bad depends on whether the purposes being served are good or bad.

For example, some group may be using the educational institution not to educate but to control, in an arbitrary way, certain minority groups. Also, a program can, and sometimes should, be evaluated not only in terms of its intended purposes, but also in terms of its unintended consequences for others. Indeed, given such consequences, one should take them into account when evaluating the educational program.

We concluded, therefore, that it is misleading, and perhaps pernicious, to say evaluation is merely an activity that provides information that plays to the major contextual setting. Such a view would misunderstand the significance of the major contextual setting and it would underemphasize, or ignore, the consequences of the program for the students and the community at large. Such an activity can rightly be censured for being partial in the extreme.

Sjoberg has argued that evaluators have as a matter of fact aligned themselves with the dominant groups in the system and meekly accept their "definitions" of program success and the public good (Sjoberg, 1975). He argued that the evaluator should adopt a "countersystem" position. Sjoberg is mistaken, however, in offering such a blanket prescription. His prescription is wise and justifiable only in cases where such "definitions" of program success and the public good are unjustifiable. Sjoberg overlooks the fact it is likely that an evaluator may well find a situation in which such "definitions" are justified. In the latter sort of case, the evaluator may rightly wish to "support" rather than counter the system.

Of course, empirical knowledge about the "major contextual setting" should be attained. But the empirical knowledge should not be used so that one can merely play to the "major contextual setting." Empirical knowledge plays a different role.

Here it is helpful to notice how social scientists view the role of their empirical findings. Paul Samuelson has written that the reason for doing economic analysis is that one's <u>increased understanding</u> of the system should aid in <u>control and improvement</u> (Samuelson, 1961, pp. 9-10). Robert Dahl has argued that political analysis helps one to understand

the world he lives in so that he might make more intelligent, more rational, and more moral choices among the alternatives. According to Dahl, political evaluation is concerned with finding which of the political alternatives is best (Dahl, 1970, pp. 1, 100). The political evaluator is not trying to have direct results; he is trying to accomplish something which will lead to improvement, directly or indirectly.

In the light of these remarks by Samuelson and Dahl, one should be troubled by the suggested division of evaluation by Ross and Cronbach. They suggested that we consider nontechnical and technical issues of evaluation and identify the technical issues with empirical research methodology matters. They seem to identify the nontechnical issues with either politics science or political analysis (Ross and Cronbach, 1976, pp. 86, 94). They have overlooked what Dahl refers to as political evaluation. Political evaluation is essentially concerned with determining whether political institutions conform to certain ethical principles. Ross and Cronbach should not leave out questions about whether a practice is ethically defensible.

Evaluation is a justifiable activity when it serves as a means of finding effective, ethical interventions. It has the task of determining whether a certain strategy is effective as well as determining whether a certain strategy is ethically permissible or obligatory. Evaluation has an important and justifiable role to play, but its effects are often more indirect, more long range.

The Sorts of Theories of Value

As we have pointed out above, to evaluate something is to find or determine the value or worth of something. The concept of value (or worthwhileness), then, is logically prior to the concept of evaluation.

A complete theory of evaluation logically depends on a theory of value. And here philosophers have an important contribution to make to the study of evaluation because of their detailed research into problems of values. Philosophers should remind us, for example, that the terms "value" and "valuation" and their cognates and compounds are used in confused and ambiguous ways, not only in economics and philosophy, but also in the social sciences and the humanities. Philosophical explication of these terms is required as a first step. For example, one should remember that the term "value" is sometimes used as an abstract noun (a) in a narrower sense to cover only that to which such terms as "good," "desirable," or "worthwhile" are properly applied, and (b) in a wider sense to cover, in addition, all kinds of rightness, obligation, virtue, beauty, truth, and holiness (Frankena, 1972, pp. 80-82).

In its widest sense, "value" is the generic noun for all kinds of critical or pro and con predicates, as opposed to descriptive ones, and is contrasted with existence or fact. In its narrower sense, "value" covers only certain kinds of critical predicates and is contrasted with descriptive predicates and even with other critical ones like rightness and obligation. The term "value" is a normative term and is to be contrasted with descriptive terms.

We can now see why it is misleading to say that the major function of evaluation is to provide information. To say that the program is a good one is not like saying that the program is a flexible one or that the program meets certain standards. In saying that the program is a good one, the person is doing something other than, or something more than, saying something true or false as one does in saying that the program is flexible, or in saying that the program meets certain standards.

Saying that Jones has received a great education is logically different from saying that Jones scored in the 98 percentile on some standardized test. An evaluation does something other than, or something more than, providing information.

The Oxford English Dictionary says among other things that the word "good" is the most general adjective of commendation, implying the existence in a high, or at least satisfactory, degree of characteristic qualities which are either admirable in themselves, or useful for some purpose. Therefore, saying something is good is not quite prescribing that we do it or saying that we ought to bring it into existence, but rather commending it, with the implication that one is doing so on certain grounds (Brandt, 1959, p. 268; Frankena, 1972, pp. 81, 106).

Philosophical investigations have also produced various theories about the nature of value. The philosophical theories of value have been of two sorts (Brandt, 1959, pp. 4-10; Frankena, 1972, pp. 4-5; Hare, 1973, pp. 39-43). Normative theories make value judgments or valuations. They tell us what is good or what has value, what is bad, and so on. They answer the questions "Which value judgments are true or valid?" and "Why?" Such theories have usually been concerned only with general value judgments. The theoretician might try to find a set of fundamental (or basic), valid principles, which is complete in the sense that all true value judgments can be deduced from it (given an adequate stock of nonvaluative or factual premises), and which is also as economical as possible, in distinct concepts and principles (Brandt, 1959, p. 5).

Metanormative theories analyze value, valuation, and good; they neither make value judgments nor tell us what is good or has value.

They define what goodness and value are and what it means to say that something is good or has value. Sometimes philosophers also offer descriptive generalizations about what is valued or regarded as good in some culture or group cultures, and explanatory theories about why this is so valued or valuable. However, this is usually ancillary to their discussions of normative or metanormative questions. Descriptive and explanatory theories belong to anthropology, psychology, and sociology.

Metanormative theories also try to answer the question or set of questions which has to do with the justification or validity of value judgments and normative theories: Can they be justified or established with any certainty by some kind of rational or scientific inquiry? Can they be shown to have objective validity in any way? If so, how? What is the logic of reasoning in such matters? Those who state that values are arbitrary--that there is no way to establish objectively the validity of certain value judgments--are making a statement which belongs in a metanormative theory.

There is much questioning about whether such a theory of value has practical use. In certain branches of the social sciences, the theory of value is a logical foundation. Robert Dahl regards political analysis as serving the task of political evaluation. One does political analysis so that one can improve the situation. Paul Samuelson holds a similar view about economic analysis. Certain economists see the task of economic theory as one of making sound proposals about the means to maximize human welfare. Political analysis and economics, then, are essentially specialized continuations of that branch of normative ethics that consists of the critique and justification of institutions.

According to Richard Brandt, some confusions of social scientists today are a result of ignorance of ethical theory. For example, he points out that some social scientists do not distinguish the question of whether any ethical statements can be justified as correct for everybody from the question of whether in fact now, as opinions stand (criticized or uncriticized, informed or uninformed), people hold different views about ethical matters. They then proceed, Brandt says, from observations of variant ethical opinions among different peoples, to a wholly unjustified inference about the impossibility of making objective assessments as to whether progress has occurred or as to the comparative merits of different institutions and customs. Hence, they conclude with a dubious theory about the logical structure of the social sciences (Brandt, 1959, p. 11).

There are writers on educational evaluation who have stated that values are essentially arbitrary and that it is quickly apparent that no adequate methodology exists for the determination of values (Stufflebeam & Guba, pp. 26, 78). Those who hold that values are essentially arbitrary are committed to holding that evaluation is no better than superstition and that there can be no justification for performing any form of evaluation. Can evaluators be nothing more than smooth-talking, modern-day sophists or modern-day apologists?

Such writers would do well to consider, for example, the work of Richard Brandt and Bernard Williams. Williams wrote that relationism, the anthropologists' heresy, is possibly the most absurd view to have been advanced even in moral philosophy (Williams, 1972, p. 20). In its vulgar and unregenerate form, he said it consists of three propositions:

1. That "right" means (can only be coherently understood as meaning) "right for a given society";
2. That "right for a given society" is to be understood in a functionalist sense; and
3. That (therefore) it is wrong for people in one society to condemn, interfere with, etc., the values of another society.

As Steven Lukes noted, "Williams easily shows the inconsistency of this composite doctrine: the third proposition uses a non-relative sense of 'wrong' not allowed for in the first proposition" (Lukes, 1974, p. 174).

Is value theory, then, of serious practical importance? The answer is clearly yes. If the conclusions of normative ethics about the ultimately worthwhile, about duties and rights, and about the critique of institutions are well established, then value theory is of the highest practical importance.

Following Brandt, we believe that the most important thing we have to gain from the tradition of ethical (value) theory is a model for a value system--what are the criteria for a tenable system of value judgments (Brandt, 1959, p. 14).

We hope it is clear, then, that anything which aspires to be a complete theory of evaluation must also concern itself with theories of value. Philosophical inquiries into the nature of value have the greatest relevance to such theory building.

Empirical Aspects

If we incorporate the philosophers' two sorts of theories, is the task of builing a tenable evaluation theory complete? The answer is no. Evaluation may have as a main purpose the criticism and justification of institutions and policies of society, and evaluation may provide us with principles and criteria to guide our conduct. Still, more is needed.

The mistaken claim that evaluation has as its prime function the providing of useful information serves to remind us of a very important point. For we need to understand our current institutions so that we can properly evaluate them on the basis of relevant facts. It may be that we need to understand how our institutions work so that we can effectively restructure them to serve our ultimate principles. At any rate, we need reliable information in order to apply intelligently our principles of criticism and justification. We need to understand the strengths and weaknesses of our actual and possible institutions and of the people who are involved in them so that our institutions can serve legitimate human values. We need to know which of the possible strategies are justifiable and will lead to an improved situation.

Such matters can be settled, in principle, by the empirical methods of the social sciences. It is a matter of fact whether a given institutional structure will serve a certain end better than another institutional structure. Empirical theories and generalizations will tell us which of certain conditions are necessary or sufficient for securing given ends.

Should we say, then, that evaluation theory, as a practical science or art consists of two branches: the normative and the descriptive branch? Or should we say that only the normative aspect deserves the title of evaluation theory?

Because of the complexity of the institutions of society, especially the educational institutions, we believe it is better to say the former: educational evaluation theory has two branches, the normative and the descriptive. And each of these branches, as we have seen, has its own kind of theories. The normative branch has the normative and the meta-normative theories, and the descriptive branch has the empirical theories.

(It should be remembered that the empirical theories also have their metatheories--namely, the philosophy of the social sciences.) Practical evaluations need to be grounded on the principles and facts of the normative and of the descriptive branch.

To have principles without grasping the facts is to be unrealistic and unable to apply properly the principles. To have facts without having principles is to be directionless. Sidney Hook has put it nicely. No wise policy can be formulated independently of the facts, but no mere recital of the facts alone determines policy (Hook, 1973, p. 113).

Conclusion

We have examined some of the major views of educational evaluation. We have offered philosophical criticism of the views about the nature of educational evaluation. We have tried to show the advantages and the disadvantages of such views.

In offering these remarks about the nature of evaluation, we have argued for the following. In its central case, to evaluate something is to determine or find the value or worth of something. Evaluation is essentially a normative activity which has as its logical foundation both ethical (or value) principles and empirical knowledge. Evaluation theory is comprised of theories of value and theories of social reality. Philosophers have an important contribution to make with regard to theories of value. The essential purpose of educational evaluation is the criticism and justification of certain social institutions. The practice of evaluation, properly understood, is of the greatest practical importance.

In closing, let us remind the reader that our work has only laid down the general outlines of a defensible theory of educational evaluation.

To have practical use, more details and concreteness must be added to the general conception. For example, given that we have correctly laid out the general conception of evaluation, how should our institutions and our evaluation teams be constructed so that the purpose of evaluation can be achieved? It should be clear that justifying an educational evaluation theory requires the combined efforts of social scientists and philosophers.

The value of our effort here is that it has laid the general, logical foundations upon which fruitful theorizing can build. In the end, a good theory is in everyone's interests. As one group of writers has recently put it:

> As it becomes clearer what epistemological stance, what view of the political system, or what assumption about the purposes of evaluation leads to each of the divergent positions taken by evaluators, researchers, and decision-makers, babble may be replaced by the kind of argument that brings the field to maturity--maturity in service, as well as of internal clarity (Ross & Cronbach, 1976, p. 106).

We must add that not only do epistemological questions need to be answered, but also ethical or value questions need to be answered. Evaluation as a practical activity is grounded on empirical facts and ethical principles.

Bibliography

Alkin, M. Evaluation: Who needs it? Who cares? Studies in Educational Evaluation, Winter 1975, 3, 201-211.

Alkin, M. The title of this speech is.... Keynote Speech, Annual Meeting of California Educational Research Association, Burlingame, California, November 18, 1976.

Baier, K. The moral point of view. New York: Random House, 1965.

Black, M. Is scientific neutrality a myth? Public Lecture, Annual Meeting of the American Association for the Advancement of Science, New York, January 1975.

Brandt, R. Ethical theory. Englewood Cliffs, N.J.: Prentice-Hall, 1959.

Campbell, D. Reforms as experiments. In M. Guttentag & E. L. Struening (Eds.), Handbook of Evaluation Research, 1975.

Dahl, R. Modern political analysis (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Frankena, W. Ethics (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Guttentag, M., & Struening, E. L. (Eds.). Handbook of evaluation research. Beverly Hills, Calif.: Sage Publications, 1975.

Hare, R. M. The language of morals. Oxford University Press, 1952.

Hare, R. M. Ethics. In Essays on the moral concepts. Berkeley: University of California Press, 1973, 39-54.

Hook, S. Education and the taming of power. La Salle, Ill.: Open Court Publications, 1973.

Lukes, S. Relativism: Cognitive and moral. The Aristotelian society, 1974, sup. vol. 48, 165-189.

Popham, W. J. Educational evaluation. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Ross, L., & Cronbach, L. J. Book Review of Handbook of evaluation research, M. Guttentag & E. L. Struening (Eds.). Proceedings of the National Academy of Education, 3, 81-107.

Samuelson, P. A. Economics: An introductory analysis (5th ed.). New York: McGraw-Hill, 1961.

Sjoberg, G. In M. Guttentag & E. L. Struening (Eds.), Handbook of evaluation research, 1975, 2, 29-51.

Stake, R. Considerations in planning evaluation studies. In B. F. Worthen & J. R. Sanders (Eds.), Educational evaluation in theory and practice, 1973, 303-316.

Stufflebeam, D., & Guba, E., et al. Educational evaluation and decision making. Itasca, Ill.: Peacock, 1971.

Tyler, L., & Klein, F., and Associates. Evaluating and choosing curriculum and instructional materials. Los Angeles: Educational Research Association, 1976.

Williams, B. Morality. New York: Harper & Row, 1972.

Worthen, B. F., & Sanders, J. R. Educational evaluation in theory and practice. Worthington, Ohio: Charles A. Jones, 1973.

FE:4

"The Foundations of Educational Evaluation"

Frederick S. Ellett, Jr.

I. Introduction

Even those interested in educational evaluation may be surprised to fully realize the scope and depth of issues and problems which fall under the rubric of educational evaluation theory. For example, a great deal of the work involves problems of statistical inference and empirical generalization; also a good deal of the work involves problems of how one can and how one should arrange the decision-making structures of a community. The topics range all the way from specific and important problems in epistemology and scientific inquiry to specific and important problems in political and ethical theory. Along the way, of course, a great many statements and claims are made which by their logical nature belong to the social sciences--to psychology, sociology, and political science. If educational evaluation theory is to be so broadly conceived, then it is important to recognize that the field of inquiry is truly interdisciplinary.

In the face of such complexity, some organization, systematization, and justification of the various positions and statements is urgently needed. Also, practice based on such positions about the nature of educational evaluation has the potential for affecting in significant ways the lives of many people in our society. Therefore, the various positions and statements, and the practices based upon them, stand in need of ethical justification. My aim here is to illustrate the very important way in which the positions and statements in the field must be

organized, systematized, and justified. That is, I shall illustrate the logical structure and the logical foundations of educational evaluation activities. Discussion and debate about educational evaluation, and the practice itself, will improve only if people become clear about the logical structure and the logical foundations of educational evaluation theories. Failure to understand the proper role of the foundations of education evaluation can only lead to confusion, error, and wrongheadedness in theory and in practice.

In this essay, I shall clarify and explicate the logical structure and the logical foundations of educational evaluation activities. In carrying out this task, I shall find it useful to talk about the various forms of educational evaluation activities, drawing a distinction between descriptive (or empirical) and normative theory. I shall illustrate the basic, logical structure of the various educational evaluation theories, explicating their logical foundations. I shall argue that claims about evaluation stand in need of justification on rational and ethical grounds.

It should be understood that I intend to leave it an open question whether a single person or a group of people with certain characteristics is required or desirable in carrying out the activities and tasks of educational evaluation.

II. Preliminary Remarks

Many people believe that it is too early in the inquiry concerning educational evaluation to decide whether it is possible to construct a single, unified evaluation theory which has such an enormous scope and range. Indeed, if such a theory is possible, it will be a theory which

is logically and structurally complex. Nonetheless, it is important in theory construction to be aware of the logical structure and the logical foundations which characterize adequate theories. The work here will contribute to theory construction and revision, and to practice, by explicating these logical features.

The careful reader, however, may be surprised to find that I will not start the inquiry with a definition of educational evaluation. In general, the use of definitions in the writings about educational evaluation has not been productive. It seems that every writer who finds a new aspect of an evaluation situation writes a definition that emphasizes his newly discovered aspect. No attempt is made to ascertain whether the definition covers all cases of educational evaluation. Writers show little or no concern for whether their definitions are too broad--so that the definition applies to things that don't usually count as evaluation--or too narrow--so that the definition fails to apply to things that usually count as evaluation. Before one can arrive at an adequate definition of terms, one needs to have a good, comprehensive set of examples of educational evaluation. Furthermore, one must investigate the adequacy of the definition with respect to borderline cases and to contrary cases and show that the definition applies to the model or central cases. Indeed, one should recognize that most writers are not interested in determining the definition of a term but are merely offering stipulations for the term.[1]

In due course, a careful determination of the definition will be quite useful. In a field that is fairly young, however, it may be, and has been, counterproductive to try to arrive at something like a definition. A definition is likely to inhibit further thinking and useful inquiry if it is done too early in the theoretical inquiry.

Fortunately, there is a way to proceed which does not require that we start with a formal definition. There are plenty of familiar cases of educational evaluation. Many of us are familiar with the evaluations of Title VII bilingual education and of Title VIII dropout prevention programs. There have been national assessments of educational progress and reassessments of the effect of family and schooling in America. Early childhood programs and "head start" programs have been evaluated. People have evaluated the use of new technologies, such as computer-based instruction. People have evaluated the graduate training programs of teacher education and the use of paraprofessionals in the schools. Alternative ways of financing (and structuring) public schools have been evaluated. People have assessed the relative merits of public and private schools. There have been many evaluations of teacher effectiveness in a school district and in a given curricular program. Various educational policies have been researched. There have been short term and long term evaluations of student achievement. Of course, the list of examples is not meant to be exhaustive. I am sure the reader can quickly add further examples to my brief list.

In general, educational evaluation is centrally interested in evaluating teaching, learning, and curriculum efficacy. But education evaluation also considers those activities, practices, and policies which are related to enhancing teaching, learning, and curriculum development. Hence it becomes concerned with the financing, administration, and, in general, the social-political organization of educational programs.

The point is this. All of us have a fairly good intuitive understanding of what counts as an educational evaluation. Until the field

of educational evaluation develops a bit further, such an intuitive understanding will serve us well.

More importantly, there are methods for becoming clearer in our understanding of educational evaluation. We should try to codify, systematize, and organize the principles, standards, and criteria, the methodolgical and procedural principles, and the aims, purposes, and goals of the activities which count as educational evaluation. It may turn out that activities which are commonly called educational evaluation can be classified in subgroups such that each subgroup has specific characteristics not shared by the others. Once this has been accomplished, we will be in a position to determine which educational activities, under what conditions, are justified. We will be in a position to develop a better understanding of educational evaluative activities. In the rest of this paper I shall set out the basic logical structure of such a codifying and organizing, and I shall explicate the logical foundations of educational evaluation activities. I intend to accomplish this by drawing the distinction between the descriptive and the normative theories and by presenting a schematic taxonomy of the various forms of educational evaluation activities.

III. Theory in the Descriptive (or Empirical) Realm

When most people think of educational evaluation theory, they are likely to think about a theory that intends to serve as a guide for practice, about a theory that is intended to lay down criteria and principles of a good evaluation. But there is another kind of educational evaluation theory which is not essentially interested in offering

guidance. It contains no normative premises for doing evaluation--rather, it is essentially interested in describing and explaining certain phenomena and occurrences in the world.

I shall refer to this kind of activity as <u>educational evaluation analysis</u>. It is important to distinguish this descriptive or empirical kind of educational evaluation activity from the other kind--the normative kind. The distinguishing characteristics of educational evaluation analysis is that it essentially aims at constructing a descriptive (or empirical) educational evaluation theory. In particular, the analyst attempts to construct a theory <u>about</u> educational evaluation activities. An analyst regards such activities as the object of empirical study and strives to find generalizations and laws which hold for the forms of evaluation being considered. His highest intention is to construct an empirical theory which would have great explanatory power so that the theory can account for the first-order general relationships.

This is not to deny that the findings of educational evaluation analysis are of importance for practical activities. Indeed, rational and moral agents necessarily try to take the relevant empirical generalizations and explanations into account when they judge the desirability of an action or a program. The point being made here is that analysis is primarily interested in finding sound and reliable generalizations and theories about the empirical characteristics of the forms of educational evaluation. The other kind of activity is concerned with determining the implications that should guide the practice of evaluation. This second kind of activity aims not at an empirical or descriptive theory but at a normative theory. A normative theory shows, at least in general outline, what is a good (bad), better, and best evaluation, and

also what is right, obligatory, permissible, forbidden, and praiseworthy. I shall deal with this kind of activity in Part IV.

The educational evaluation analyst, then, is essentially interested in the scientific understanding of educational evaluation. He is not directly interested in actually doing educational evaluation. The educational evaluation analyst is similar to Robert Dahl's political analyst. As Dahl points out, skill in analyzing politics is not the same as skill in practicing politics. Of course, every skilled, political practioner must have some capacity for political analysis, even though he may be unable to explain or to formulate what he knows. The political analysts sees it as his task to set out to scrupulously test his hypotheses, generalizations, and theories against the evidence of experience by means of meticulous observation, classification, and measurement, and so on. He strives to obtain scientific understanding of the political system.[2] Also, the evaluation analyst is similar to Paul Samuelson's economics analyst whose primary goal is to describe, to analyze, to explain, and to correlate the activities of production, unemployment, prices, and similar phenomena. To be significant, such descriptions must be more than a series of disconnected narratives--they must be fitted into a systematic pattern, into a theory.[3] According to Samuelson, economics is not business management. It does not tell one how to make a million dollars, to prepare annual accounts, or to anticipate the stock market.[4]

An educational evaluation analyst, then, sets out to formulate a systematically related set of statements, including some lawlike generalizations, that is empirically testable. He sets out to discover empirical generalizations that govern the phenomena, and he hopes to construct a

theory that in some sense explains the various generalizations.[5] The analyst is essentially interested in theory construction, confirmation, and refutation. It is important to notice here then, that the analyst ultimately intends his work to meet the standard criteria of scientific inquiry and scientific understanding.

In the beginning, of course, the analyst may be well advised to look not only at the various evaluation activities but also at activities which are concerned with accounting, appraising, assessing, auditing, budget managing, determining effectiveness, monitoring, and so on. Although these activities may go by different names, it might turn out that the activities are similar to or the same as certain forms of evaluation. Such a comparative study is likely to bring out some important similarities and differences which will help the analyst to better understand the various forms of evaluation activities. The analyst shouldn't draw his investigative sights too narrow until he has a rather clear idea of the terrain of the objects he seeks to understand.

It may be helpful here to consider some examples of the kind of questions which analysis sets out to answer in its activities. The answers to these quesions require empirical verification. The following are to be typical and suggestive illustrations of these questions. What is it that people are doing when they say they are evaluating some educational object? What are the kinds or forms of evaluation? What standards, principles, and criteria are being used, implicitly or explicitly, in evaluating an educational object? What effect do the various forms of evaluation have on educational decision making? What effect could these forms have if certain institutions were restructured? What effects do the various forms of evaluation have on the public's

perception of education and educational politics? In what way is evaluation a political process? Does evaluation help to perpetuate the false view that education is apolitical? Who controls the various forms of evaluation? What other possibilities exist for managing evaluation in our society? Who takes part in the forms of evaluation? What are the possibilities for other people or other groups taking part in the evaluation process? What effects do the various governmental bodies and levels have on the evaluation process? In what ways do federal and state statutes help (or hinder) evaluation? What possibilities exist for restructuring federal and state statutes to enhance evaluation activities? What kinds of people are needed to do evaluation? Does (can) evaluation (or certain forms of it) help us in achieving rational (and just) allocations of educational resources?[6]

To further illustrate the point, I have selected some assertions which have been put forward in the literature, but I have written them as questions because they seem to stand in need of defense. One should notice that, in general, writers are not very careful in delimiting which kind or form of evaluation they are talking about. Before one can get an empirical answer to these questions, one needs a clearer and more precise understanding of what counts as evaluation for each.

Here are the examples. Given the social demand for evaluation, how can one account for the fact that relatively little genuine evaluation is going on at present (Worthen & Sanders)?[7] By the way, is there a social demand for certain forms of evaluation? Which groups in society are interested in what kinds or forms of evaluation? Is the evaluator likely to provide less ammunition for advocates of social intervention than for conservative critics of social intervention (Carol Weiss)?[8] Do

evaluators, in certain situations, align themselves with the dominant groups in the system and meekly accept their definitions of program success and the public good? (Robert Stake)[9] Are evaluations only used by those whose preconceptions and values match with the evaluations' results? (Carol Weiss)[10] Do all evaluators want to improve education? (W. James Popham)[11] What forms of evaluation hold great promise in providing educators with badly needed information which can be used to improve the process of education? (Worthen & Sanders)[12] What are some ways which can be used to improve the technical quality of evaluation research? (Cook & Grunder)[13] If one is interested in solving educational problems, how much can evaluation contribute in comparison with what basic research into the social processes and institutional structures might contribute? (Carol Weiss)[14] What institutional and political arrangements are required to do effective evaluations? How can such institutions and arrangements be created? (David K. Cohen)[15] In what ways are evaluations conditioned by the political context? (Michael Patton)[16] How do political factors emerge to affect the practice of evaluation? (Patton)[17] Are the barriers to evaluation simply another manifestation of the obstacles to federally initiated reform when most power is local and when reform is a relatively low national priority? (David Cohen)[18] Does Cohen's statement that reform is a low national priority contradict Worthen's and Sander's statement that there is a social demand for evaluation?

Given an understanding of the nature of this kind of question, one should realize that every essay on educational evaluation either makes certain empirical statements or presupposes certain empirical statements. The writer may say that performing certain activities in a certain

situation will result in gathering information of a certain kind. The writer may say that a good evaluator has to perform certain activities, but it may be unclear whether anyone could perform the activities. Most often such statements stand in need of verification or scientific test. The verification of such empirical statements about educational evaluation is the task of the educational evaluation analyst.

It can be seen, then, that educational evaluation analysis strives to be a social science that studies a particular set of phenomena--educational evaluation activities. It may well be that such a social science needs to be interdisciplinary, drawing upon the talents and strengths of the sociologist, psychologist, economist, and political analyst, among others. The major point here, however, is that the activities, the inquiries, and the conclusions of the educational evaluation analyst essentially aim to meet the standards of scientific excellence in the wide sense. They aim to meet the standards of inductive and statistical excellence, the standards of empirical science. Such inquiry and theorizing belongs in the empirical realm.

IV. Theory in the Normative Realm

A. A General Characterization of the Normative Realm

In Part III, I considered one kind of activity which is concerned with educational evaluation. I argued that the distinctive feature of educational evaluation analysis is that it is essentially an empirical (or descriptive) activity that aims at finding a systematically related set of statements about educational evaluation, including some lawlike generalizations, that is empirically verifiable. It is ultimately

concerned with constructing a theory about the forms of education evaluation and their empirical possibilities and capabilities for bringing about certain things. Although any justifiable recommendation for the practice of evaluation must be based upon the empirical findings of analysis, the analyst, as an analyst, is not concerned with, nor entitled to make, recommendations for the actual practice of evaluation. Additional nonempirical premises are logically required in order to draw recommendations from the empirical analysis. Such premises include principles, criteria, and standards which belong not to the realm of the empirical analyst but to the realm of the normative theorist.

In contrast to the activities of analysis which are empirical in nature, the premises, principles, conceptions, and theories which I will consider in this section are normative. Although the theories may assume or depend in certain ways upon factual (or empirical) statements, they are normative because they tell us what evaluation activities are good or bad, right or wrong, adequate or inadequate, rational or irrational, just or unjust. They implicitly claim that evaluation activities ought (or should) be conducted in certain ways; they talk about the evaluator's obligations, responsibilities, and duties. Some offer recommendations about how evaluations should be done in order to promote rational and just decision making. In general, they set out to give advice, recommendations, warnings, guidance; they offer maxims, priority rules, strategies, and decision rules; they point out pitfalls, problems, demands, and restrictions. These activities are normative, then, because they make claims about whether certain evaluation activities are, for example, good or bad, right or wrong, sensible or not, rational or irrational, just or unjust, obligatory, permitted or forbidden. They

are normative because they implicitly or explicitly include, and often stress, principles, criteria, and standards by which evaluation activities are to be distinguished as, for example, good or bad, right or wrong, sensible or not sensible, adequate or inadequate, rational or irrational, just or unjust.

When people engage in such normative theorizing, they are doing something other than describing evaluation activities. They are doing something other than offering true, empirical generalizations about evaluation activities. A person engaging in such normative activities is implying that certain standards, criteria, and principles ought to govern evaluation activities; he is implying that such standards, criteria and principles are the proper or appropriate ones to govern the activities of the evaluator. Such theorists are giving prescriptions for conducting evaluations.

Almost everyone who has written about educational evaluation has made one form or other of such normative claims. Some have spent more time than others in clarifying, canonizing, and systematizing the standards, criteria, and principles that are meant to characterize good evaluations. A few have tried to defend or to justify their normative claims. Unfortunately, much work remains to be done.

In this section of the paper I will show that there are (at least) five different kinds or forms of normative theories of educational evaluation. I have not tried to develop a rigorous and complete taxonomy which locates and places every form of normative evaluation theory. My work here will provide a necessary step toward the systematizing and completion of the set of standards, criteria, and principles which govern each of the kinds or forms of evaluation activities. Eventually,

one should try to find a complete set which is economical in the sense that it contains those distinct concepts, standards, criteria, and principles from which all the other normative claims can be deduced.[19] Of course, a complete accounting of a theory will need to make explicit the empirical assumptions and generalizations which form an essential part of any normative theory of evaluation. Any normative theory of evaluation necessarily includes a view about the facts of human nature and society and especially about the possibilities for education. Upon such premises the theory builds its recommendations about the evaluator's immediate aims, knowledge, methods, responsibilities, and duties which are seen as means to carrying out the global purpose of evaluation.

Therefore, I present this tentative and schematic taxonomy in order to underscore a very important point. Most evaluation theories are normative and their principles, criteria, and standards need to be systematized and organized into a complete set which is economical in some sense.

In the following subsections I have also tried to clarify the kind or form of evaluation activity by suggesting that various writers on evaluation in some ways typify that kind or form of activity. I believe this strategy is quite useful as long as one remembers that the categorizations are somewhat rough and that many writers are often engaged in more than one kind or form of activity in a single essay. Clarification and advancement will occur, however, only if one organizes the various claims of these writers into their appropriate kind of activity.

As I shall show, the positions and views of the writers differ in the principles, criteria, and standards which define their conception of the evaluator's responsibilities and duties. Unless this diversity of

principles, criteria, and standards is understood, much of the discussion about evaluation is likely to be unproductive because the people will fail to realize that they are talking about different things.

One further comment, it should not be assumed that all forms of evaluation goodness are implicitly or explicitly related to rational or to moral goodness. It must be recognized that many of the most popular normative evaluation theories have no essential connection with rational goodness or with moral goodness. A few of the normative evaluation theories recommend activities which are logically incompatible with rational goodness or with moral goodness. Normative evaluation theories present a view about what is in some sense good evaluation and about what are the responsibilities, obligations, and duties of an evaluator. One must not assume, which in most cases would be false, that the goodness and the responsibility are essentially connected with rational and moral goodness and responsibility.

B. Five Forms of Normative, Educational Evaluation Theory

As I have said earlier, a normative, educational evaluation theory is a theory which says what a good evaluation is and what the responsibilities, restrictions, and obligations of an evaluator are. It recommends how an evaluator ought to conduct himself in performing his activities. The principles, criteria, and standards of the theory determine what a good evaluation is and what the responsibilities, restrictions, and obligations of an evaluator are according to that theory. Although all of the theories are normative, notice that the principles, criteria, and standards do differ in important ways from one theory to another. Thus, there are cases where what is a good evaluation for one theorist is not what is a good evaluation for another. While there may be some principles,

criteria, and standards which are common to all normative theories, such commonalities are of a relatively small significance. This point will become clear as we consider the various forms of evaluation theories.

1. Empirical Hypothesis Testing and the Ultimate Principle of Evaluation

1a. A good evaluation is a good social experiment. I will begin with one of the major positions in the field. This kind of educational evaluation theorist holds that a good evaluation is essentially a good, social experiment. In other words, the theorist claims that an evaluator should conduct social experiments. This would include manipulating the independent variables and measuring the dependent variables to determine scientifically whether the program causes certain results. Such theorists hold as fundamental the principle that one should follow the methods of experimental inquiry, with random assignment of subjects to an experimental group that is exposed to the program stimulus or to a control group. David Cohen, Marcia Guttentag and Elmer L. Struening, and Carol Weiss appear to hold such a theory of educational evaluation.[20]

The theorist also holds that the evaluator should communicate the findings in a manner that will facilitate decision making, but he holds that the evaluator should try to arrange the project so that it facilitates the experimental appraisal of the program's effects. In some cases, however, it may not be possible, for various reasons, to follow the methods of experimental research. In these cases, there is disagreement among the theorists about what the evaluator should do. Some say he should not get involved; some say he should go to a "weaker" strategy; some say that he should work to change the institutional structure so that the experimental methods can be employed.

What characterizes the position, then, is the claim that in general the good evaluator is the one who employs the experimental methods of inquiry. He should follow those procedures and strategies which enable one to employ the experimental method.

In addition, it is important to notice that such a theorist also holds that the following principle should govern the evaluator's activity. The evaluator should determine whether the program satisfies the objectives and expectations of some specified decision maker or resource provider. Except for clarifying and defining his objectives and expectations, the evaluator does not need to have any other social, personal, or political skills. The evaluator should mainly possess the skills and knowledge of an experimental researcher.

1b. A good evaluation is a good empirical inquiry. A variation of the them of the researching evaluator is the view that the evaluator need not do experimental inquiry but that he should do the next best thing, say performing quasi-experiments, perhaps using causal-modeling procedures and multiple regression techniques. As the previous theorist holds, this theorist would hold that the evaluator should take as given the objectives and program definition which some specified decision maker or resource provider has provided. According to the theory, the evaluator is not responsible for determining whether the program has unintended consequences which are bad or undesirable. The evaluator is only responsible for determining to what degree the program achieves its intended objectives.

Thomas D. Cook and Charles L. Gruder, in a recent journal of education, hold the more general theory that evaluation is essentially empirical hypothesis-testing, that it aims at testing (and verifying) causal

propositions about policy-relevant relationships.[21] The Encyclopedia of Education Evaluation, edited by Scarvia B. Anderson, Samuel Ball, Richard T. Murphy, and Associates, shares the general theory that evaluation bases its methodology on the social sciences.[22] For them, evaluation is empirical hypothesis-testing where its canons of excellence are those of social science methodology. Notice that since the justification of the ultimate principles of rational and of moral (or ethical) conduct appear to be out of the reach of scientific methodology, crucial assumptions and premises of evaluation practices will not receive their proper criticism and scrutiny. The elucidation of the "mainstream theory" of evaluation given by Cronbach and Ross is quite similar to theories of evaluation considered in this section.[23]

The principles that forbid or discourage the evaluator from evaluating the worth of the objectives or the desirability of unintended consequences of the program seem to have a scientific-research flavor about them. The theorist is likely to believe that a scientist is an empirical researcher. As such he is in no position to make judgments about the worth of the ends of a program or about the desirability of unintended consequences. The theorist is likely to believe that the researching evaluator should be responsible only for determining the degree to which the program brings about its intended objectives. Such a determination is essentially an empirical matter of determining the causal efficacy of a given program. The researching evaluator should be responsible only for making such causal or empirical judgments. For other kinds of judgments the researching evaluator is not responsible.

2. Organizing the Principles of Measurement and Epistomology

A second form of evaluation theory, which is closely related to the first, sets out to clarify, prioritize, and systematize the standards, criteria, and principles, of empirical inquiry and inference. Such a theorist holds that an evaluator should have certain statistical knowledge and research skills. Thus, the evaluator should know about measurement, statistical inference procedures, criterion-referenced and norm-referenced testing, design procedures, sampling procedures, and analyzing and reporting procedures and strategies. The evaluator should be able, in some sense, to apply such knowledge and know how to practical situations.

For such a theorist, then, an evaluator should have knowledge of various procedures and strategies of research methodology and statistical inference. The primary obligation of an evaluator is to bring this expertise to bear in assessing the program.

Perhaps an example of this kind of educational evaluation theory is W. James Popham's Educational Evaluation.[24] Like many other books on educational evaluation, the book spends almost all of its time talking about the measurement, statistical, and methodological issues. Of course, many writers either assume as proper or merely assert to be proper, certain epistemological and methodological procedures and strategies. It is important, therefore, that a writer take care to clarify and systematize the procedures and strategies which he implies are centrally related to doing good evaluations. Writers should be encouraged to clarify and systematize other methodological presuppositions and assumptions on these matters. It is important to know where important differences lie and what kinds of issues are involved.

According to this theory, then, a good evaluator is one who has knowledge of various research methods and strategies and of various principles of statistical inference. There is an ideal version and a practical version of this form of theory. The ideal version attempts to codify and systematize the various empirical strategies, procedures, and principles which would be applicable under ideal conditions. The practical version attempts to codify and systematize which would be applicable under real life conditions.

It is important to realize that in general such theories downplay or fail to take into account the social, political, and ethical aspects of an educational evaluation.

3. Organizing the Principles of a General Approach

A normative evaluation theory need not concern itself with the rather technical issues of epistemological and methodological matters. A theory can try to develop a general structure of the recommendations which are meant to guide practice. For example, a theorist might argue that the recommendations for evaluating curricular and instructional materials can usefully be grouped into the following general categories: (1) Rationale, (2) Specifications, (3) Appropriateness, (4) Effectiveness, (5) Conditions, (6) Practicality, and (7) Dissemination.[25] Such a theorist holds that it is desirable that when the curriculum materials are developed they should be accompanied by a manual that makes every reasonable effort to describe which of the seven components have been taken into account.[26] The theorist holds that the purpose for clarifying and developing the general structure is to contribute to a greater understanding of what constitutes quality materials, quality programs.[27] Notice, however, that the general structure developed by Tyler and Klein

tends to be rather abstract and formal in its recommendations. For example, they urge that the value of the objectives must be substantiated (R_1) and that the "source" from which the objectives are derived must be indicated (R_2). Little time is spent, however, in explaining or recommending precisely how one is to go about carrying out these tasks.

To a large degree, such a theorist tends to overlook (or to underemphasize) the practical problems which are connected with carrying out the recommendations. As the theorist (form 2) who strives to clarify and systematize the epistemological and methodological procedures does, this theorist is apt to strike the would-be practitioner as being idealistic in a certain sense. For few specific guidelines or recommendations are offered for dealing with the social, political, and ethical aspects of the educational evaluation.

4. Providing Information Useful to the Decision Maker

A fourth kind of evaluation theory is one which has gained support from many of the more noted evaluation theorists. It is one which often lies behind actual practice. The normative theory which governs this set of activities has as its foundation principle the following: The evaluator should provide information which is useful for the decision maker. Of course, without further specification and clarification of the terms "information," "useful," and "decision maker," this principle is extremely vague and ambiguous. As one might expect, then, within this general form of evaluation activities there are important differences between the types of evaluation activity which a particular theorist will recommend.

4a. Counsulting evaluation. One theorist holds that the evaluator should be involved in the assessment of the educational program throughout

the year, provide information to help teachers and administrators get the program onto its intended course, to help keep the program on course, or to help the managers and teachers redirect the program so that it will achieve a desirable outcome. The evaluator should regard himself as a member of an educational team, the member who has special expertise in collecting and analyzing data about the program. He should take into account the informal, as well as the formal, aspects of relationship with others, paying particular interest to social and personal relationships. Of course, he should also prepare (and interpret) yearly evaluations of the program's status. His responsibilities also include making sure the state and federal statutes regarding evaluation are being satisfied. The evaluator, however, should strive to develop and maintain various strategies so that the information he gathers is maximally relevant to the ongoing and emerging problems of the program directors and teachers. The evaluator should be prepared to employ techniques and strategies other than standardized tests and experimental inquiry. His approach to the evaluation should be flexible and dynamic. Because of the rather informal and personal nature of the ongoing relationship between the evaluator and the program director and teachers, it might be fitting to call this consultative evaluation.

Recent work by Alkin, Daillak, and White, and by Patton seem to be good examples of what a consultative evaluator looks like.[28] It is interesting to notice that the "alternative and extended view of evaluation research" suggested by Cronbach and Ross seems to be quite compatible with the theory of the consultative evaluator, especially if they were to emphasize the role of personal factors in performing good evaluations.[29]

4b. Accounting evaluator. A related but distinct view holds that the evaluator should not be as directly or personally involved in the running of the educational programs as the consultative evaluator. The evaluator should furnish information which is most useful to the decision maker in the following way. The evaluator should determine through interaction with the decision maker (or client) what are the alternatives the decision maker will consider. Furthermore, the evaluator should determine, again through interaction with the client, what are the values or criteria the decision maker will apply in judging the alternatives. Having determined these things, the evaluator goes out and collects information which is relevant to making a decision among the alternatives with respect to the specified criteria. In general, the evaluator should be a rather disinterested, neutral professional who mainly applies his technical expertise in measurement and research methodology to the problem situation as defined by the specified decision makers.

Within this general type there are a variety of positions depending upon how one determines the values or criteria. Each theory would specify whether or not the expectations of the board of education, the expectations of parents, the expectations of the decision maker's coworkers, and the personal (or private) values of the decision maker should be considered, and if so, how much weight these various expectations should be given. It is possible, although no one has done it, that such a theorist might recommend that a specific, concrete procedure should be used to determine the values or criteria.

Given that the evaluator should consider himself as a disinterested and neutral professional who merely applies his technical expertise and

draws and reports the empirical results, it seems fitting to call such activities accounting evaluation. The views of Marvin Alkin in the late 1960s are quite similar to such a theory.[30] Anderson, Ball, Murphy, and Associates; Robert Stake; and Stufflebeam and Guba have expressed views that are closely related to accounting evaluation.[31]

4c. Expedient evaluation. A view related to the previous two, but distinct, holds as the others do, that the evaluator should provide the decision maker with any information the decision maker thinks is important. The evaluator should not criticize the decision maker but should help him get that information which will help him achieve his goals. The view holds, however, that the evaluator should do such things as allowing the decision maker to write (or rewrite) the evaluation report so that the information presented in the report best serves the decision maker's purposes, even when the empirical data does not support the report. Such activities are fittingly called expedient evaluation. (Machiavelli would have been proud of such an evaluator.) I hesitate to say that anyone has openly expoused the theory of the expedient evaluator. Although I have seen papers in the literature, I hesitate even to call attention to persons whose descriptions of their activities have a great similarity to expedient evaluation. It is my conjecture, however, that if empirical data were available to characterize the actual activities of evaluators, one would find that a good deal of it closely resembles expedient evaluation.

5. Rational and Ethical Evaluation

A fifth form of normative, educational evaluation theory strives to construct a different kind of general theory which will serve as a guide for doing good evaluations. Here the theorist is not content to have

the empirical findings of the educational evaluation analyst, nor is he content to have formulations of principles which are too remote or abstract, for various reasons, to be of use in guiding behavior. The theorist strives to develop a set of principles which will enable the evaluator to conduct a good evaluation in a given situation. The theory would provide the practitioner with practical ways and strategies for achieving the global goals of evaluation. Although the same methods and strategies need not be applicable to all situations, the theory would provide some methods or strategy for all situations.

This kind of evaluation theory has an additional and quite important feature. The theorist is essentially interested in finding those methods and strategies that can be justified on rational and ethical grounds. For such a theorist, educational evaluation has as its major purpose the criticism and justification of the various forms of teaching and learning within society and of those activities and institutions which are related to education. The criticism and the justification of the various activities are based on the principles of rational and ethical conduct--that is, those principles that should in a strong sense govern conduct. The evaluator should try to determine whether a given form of teaching and learning in a certain situation has rational and ethical value. In certain situations the evaluator should try to determine whether the related practices and policies have value that can be defended on rational and ethical grounds.

Ultimately such a theorist would like to develop a comprehensive and systematic theory which would provide practical guidance in all of the particular situations. He would like to develop a theory which would tell one what a reasonable strategy would be in certain situations.

In the end, such a theory of evaluation would have close relations to a general theory of practical reasoning. After all, in our everyday lives when we try to decide by reasoning whether a course is good (or bad), right (or wrong), or worthwhile (or not worthwhile), we are necessarily trying to evaluate the possible courses of action on rational and ethical grounds. Perhaps it might be fitting to call it practical reasoning evaluation.

Such a theorist emphatically rejects the view that an evaluator should only be concerned with determining the degree to which the program achieves its intended objectives. The theorist rejects the view that an evaluator should only consider whether the means lead to a given end and not consider the worthwhileness of the end itself. Thus, according to this view, the evaluator should be concerned about the worthwhileness of the objectives of programs. He would also claim that the evaluator should also be concerned about the desirability of the unintended consequences of a program.

Although almost all evaluation theorists hold the view that the activities of evaluation are somehow connected with criticizing and justifying actions on the grounds of rationality, one must regard such comments as ritual obeisance to rationality. For little or no effort is made to determine what the relevant principles of rationality are or to show that their theories of evaluation lead to more rational practices.

It is most important to recognize the following. Almost all of the evaluation theorists seem to be unaware of the crucial difference between rationality and morality. The principles and criteria of rational conduct are logically distinct from the principles and criteria of moral conduct. A policy could be rational and yet not be moral. I hope that

the reader grasps the potentially harmful consequences of not fully understanding the logical difference between rationality and morality.

There is not adequate space here to develop and defend the claim that there is a significant logical difference between rationality and morality. In a rough way, in terms of their applications to actions, the distinctions may be put thus: questions of rationality involve the agent's aims and the best way to realize them, whereas questions of morality involve the assessment of the pursuit of own's own aims in the light of the claims of others. A rational action is one which is in the agent's self-interest. Rationality is closely related to prudence. A rational action need not be a moral action. A moral action is one which takens into account the interests of all the individuals involved and assigns equal weight to them all. Morality is closely related to beneficence and justice. Moral questions--questions of what should we do--have absolute priority. Nobody has the right in general to neglect relevant moral considerations.[32]

That almost all theorists have failed to show how their theories can be justified on moral grounds is a fact which is likely to have pernicious consequences for students and for the general public.

Out of the better known evaluation theorists, Michael Scriven holds a position that comes closest to the position briefly described here. For Scriven seems to be one of the few to urge that the goals of educational programs should be evaluated. He holds that various people in education cannot be regarded as immune from criticism in setting up their goals.[33] He says we have an obligation to evaluate the goals as well as the means. The only plausible way of understanding Scriven's remarks here is to say that such an obligation is based on both the

principles of rationality and the principles of morality. Given his remarks about values and morality, however, it should be clear that Scriven is a considerable distance from the position of rational-ethical evaluation described in this section. His account of values, and especially morality, seems to be largely mistaken.[34]

Worthen and Sanders also hold that the goals of educational programs should be evaluated. Their suggestions that values can be equated with preference, which they say has precendent in economics, presnts another view of moral value wich must be rejected as unsound.[35]

Nonetheless, such work is a step in the direction of an adequate theory of rational and ethical evaluation. Much more work needs to be done to formulate adequately the form of evaluation activities which can be justified on rational and ethical grounds.

In closing this section, let me remind you that the taxonomy of the forms of normative evaluation theories is tentative and schematic. My aim is to illustrate a potentially useful strategy, rather than to apply it in boring detail. I have only taken a few steps toward developing a comprehensive taxonomy which could encompass all normative evaluation theories. By organizing, codifying, and systematizing the various forms of evaluation in the proper ways, the taxonomy allows one to see the logical structure of evaluation theories more clearly. It should be clear that normative educational evaluation theories do differ with respect to primary goals and intentions, methods of approach, and standards of excellence, and that they do differ with respect to obligations, responsibilities, and duties.

V. The Foundations of Theories of Educational Evaluation

I am finally in position to make the second, major point for which this essay was written. I want to draw attention to the logical foundations of educational evaluation theories. Understanding the foundations is important. To be confused about the foundations of any theorizing is to leave the theorizing open to confusion and error. It is only upon these foundations that sound and defensible theories can be built. Once those interested in educational evaluation grasp the importance of the foundations in their theorizing. Their argumentation and discussion will lead the field toward greater comprehensiveness, productivity, and defensibility.

As I have set out in the two previous sections, there are various kinds and forms of educational evaluation theories. I have also pointed out that it is important to draw the distinction between those theories which are essentially descriptive (or empirical) and those theories which are essentially normative. I have referred to these as the empirical realm and the normative realm of educational evaluation theory. Two major quesions need to be answered concerning each form of evaluation theory in each of the realms.

First, to what degree is it possible to succeed in the activities of a given evaluation theory? For example, to what extent is it possible to succeed in discovering empirical generalizations and explanations that the educational evaluation analyst sets out to find? Again, to what extent is it possible to succeed in evaluating and educational program if one uses only the experimental approach? These questions ask about the empirical possibilities of achieving the primary goals of

certain evaluation theories. Such questions ask whether it is in fact possible, and to what extent, to carry out these activities that are central to the theory.

Second, why should anyone be allowed or obligated to engage (or to support) in the activities of a given evaluation theory? Is it worthwhile for someone to engage in those activies central to a given theory? Should we value the activities of a given theory? These questions ask whether it is rational and ethical to try to achieve the primary goals of a given evaluation theory.

I will begin with some remarks about educational evaluation analysis. Of course, we already know some rather trivial empirical generalizations about evaluation. However, there is much we still need to learn about the various forms of educational evaluation. (I gave many examples of some of the important questions in Part III.) Furthermore, discussions of educational evaluation implicitly or explicitly assume that certain empirical claims are justified. Most of those empirical assumptions stand in need of empirical justification. A reasonable discussion of education evaluation must be based upon those empirical statements which can pass the empirical test. Therefore, a good reason for supporting and encouraging the work of the educational evaluation analyst is that such work will probably provide a firm, empirical foundation for any reasonable discussion of educational evaluation. In order to plan evaluation programs and strategies in a reasonable way, one needs to know the empirical (or causal) properties of the various forms of evaluation. One needs to know under what conditions it is likely that a certain form of evaluation will produce certain results. The work of the educational evaluation analyst is most likely to provide the firm foundation for such empirical statements.

Thus, when asked why one should be allowed or obligated to engage in a certain kind of educational evaluation activity, one might respond by arguing that the work of the educational evaluation analyst is needed so that we might better understand and know about the various forms of evaluation. But many people would rightly argue that the basic reason for acquiring such understanding and knowledge is essentially practical. We must strive to secure the understanding and knowledge in order to build evaluation strategies and programs that are rational and ethical. That is to say, the basic or foundational principles upon which any theory of educational evaluation must be based are the principles of rational and ethical reasoning. The findings of the educational evaluation analyst are relevant to the critique and justification of institutions, including those institutions and programs that are involved with various forms of evaluation. In the end, the analyst provides the knowledge which will help us develop institutions that are sensible and will serve human welfare.

Thus, educational evaluation analysis, an activity which is essentially concerned with discovering the empirical characteristics of the various forms of evaluation, is worth doing in so far as it contributes to the criticism and justification of educational programs on the basis of rational and ethical grounds. Educational evaluation analysis is ultimately founded on rational and ethical grounds.

It is importnat to notice that this view about the foundations of educational evaluation analysis is similar to views about the foundations of economics and of political analysis. For example, Paul Samuelson holds the reason for doing economic analysis is that one's increased understanding should ultimately aid in control and improvement of economic

policy. Although the basic questions concerning right and wrong economic goals cannot be settled by science as such but by ethics, economics can contribute to the final solution by pointing out the feasible alternatives and the true costs that may be involved in the different decisions.[36] In a similar way, Robert Dahl holds that the best reason for improving one's skill in political analysis is that it helps one to understand the world he lives in, to make more intelligent choices among the alternatives he faces and to influence the changes. Ultimately, such activities are founded on ethical and rational grounds.[37]

Within educational evaluation analysis, of course, there are some important and unresolved questions. For example, it may well be the case that educational evaluation analysis should explicitly use value terms--such as justice and rationality--in its theoretical vocabulary. For it is an empirical question, and a very important one, whether certain forms of evaluation lead to or help produce just and rational results. In other words, the analyst may need to employ such value terms in order to investigate the advantages and disadvantages of the various forms of evaluation. Clearly, part of what we mean by an advantage is that it brings about a more rational or just situation. Whether the educational evaluation analysis should employ such value terms is a question that deserves serious discussion and debate by the analysis.

So far I have argued that, properly understood, educational evaluation analysis is founded on empirical grounds and on rational-ethical grounds. When one considers the various forms of normative evaluation theory, one must come to similar conclusions about their foundations.

First, any normative evaluation theory assumes that certain empirical statements hold. In recommending a certain course of action, the theory

assumes that the course of action is indeed open for one to take and that the course of action is likely to bring about the intended results. Both of these assumptions are empirical. They assume that certain things hold as a matter of fact. Clearly, then, a reasonable, normative evaluation theory must be based upon those empirical assumptions which can pass empirical test. For example, one should know under what conditions the guidelines and recommendations of the normative theory are most likely to bring about the intended result. Any justifiable normative evaluation theory must be founded, in part, on the warranted statements of educational evaluation analysis. Ultimately, a justifiable normative evaluation theory is founded on empirical grounds.

A normative evaluation theory, then, presupposes and assumes certain empirical statements. No normative theory can be justified if such empirical statements are unwarranted. But there are also the normative premises that play a determining role in the theory. A normative evaluation theory makes certain recommendations and offers certain guidelines and maxims for practice. It implies that it is good (or right) to follow certain principles; it holds that the evaluation has certain responsibilities, obligations, and duties. Here it is important to ask whether one is permitted or obligated to follow the guidelines or the recommendations of a given normative theory? Is one permitted or obligated to engage in activities which fulfill the theory's responsibilities, obligations, and duties? Is it worthwhile for people to engage in (or to support) these activities? In order to show that the normative theory is justified, one must show that the normative premises are justified. One must show that, in certain situations at least, people are justified in engaging in the activities prescribed by the theory.

And one must determine that the theory is justified by showing that it be founded on rational and ethical grounds. Roughly speaking, it must be shown that engaging in the activities of the theory is reasonable in a certain sense and that it promotes justice and human welfare. Given that the activities of a certain form of evaluation have great potential for harm and wrongdoing, an especially strong justification may be required. One is permitted or obligated to carry out the activities of a given normative theory only if it can be shown that performing the activities can be rationally and ethically justified.

Thus, a normative evaluation theory, with its principles, standards, and criteria, and its recommendations, guidelines, and related tasks, must be founded on the principles of rational and ethical reasoning. A particular normative evaluation theory is justifiable in so far as it contributes to the criticism and the justification of educational programs on the basis of rational and ethical grounds. Normative evaluation theories are ultimately founded, in part, on rational and ethical grounds.

It should not be assumed that only a single, normative evaluation theory can survive the empirical test and the rational-ethical test. Quite the contrary. It seems plausible to hold the position that several normative evaluation theories are justifiable. It seems plausible to hold the position that one normative theory will be most appropriate under certain conditions, while another normative theory will be most appropriate under different conditions. It may also be plausible to hold the position that some of the normative evaluation theories are compatible with one another and that they should be united to create a new and more comprehensive normative theory. It may also prove useful

to build upon the insights and strengths of the various normative theories currently under consideration by taking only certain principles and strategies from each normative theory in order to construct a more justifiable and comprehensive theory.

At the present time much discussion and argumentation is needed to assess the relative merits of various normative theories. It would be wiser not to try to determine whether one of the theories is unilaterally defensible but to try to determine under what conditions the various theories are likely to bring about what results. For each of the normative theories I have considered above has prima facie difficulties due either to its limited range of application or to its limited comprehensiveness.

For example, in order to offer realistic advice a normative theory which focuses primarily on measurement, methodology, and induction must be supplemented by a theory which adequately deals with the social and political aspects of the situation. Furthermore, almost all of the normative evaluation theories, and certainly the dominant ones, take as given and not subject to review, the intentions, goals, and values of some decision maker or group of people. Some have taken the preferences of various groups as the basis for determining the appropriate values for judging the program. What are the likely consequences of such an activity? Can such activities be unconditionally justified on rational and ethical grounds? Many theories presuppose that certain ways a community makes its decisions about educational programs are justifiable. The theorists often disagree about the way in which the evaluator is supposed to deal with the decision making body. In making such statements, a theorist is presupposing or assuming an important position in social and political philosophy. In a democracy, community decision

making must satisfy certain principles. Do the theories of evaluation maintain positions that are incompatible with democratic principles?

All of the normative theories described above need to give serious attention to their view about whether the theory is involved in evaluating schooling or in evaluating education. For the great thinkers, education was far more than mere schooling. They held that education should initiate people into the highest forms of knowledge and into the most worthwhile forms of life. Central to their view was the role of critical thought. What are the advantages (and disadvantages) of an evaluation theory which is primarily interested in evaluating schooling? Should such a theory have our complete support? Will the theory need to be supplemented by additional principles so that one could evaluate the goals and outcomes of schooling?

With respect to procedural questions, we need further empirical evidence about the advantages (and disadvantages) of flexible, informal, and dynamic strategies in comparison with rigid, formal, and static strategies of evaluation. For example, can the evaluator who relies on rather informal sources of information reach causal conclusions which are as sound as the causal conclusions reached by the large scale, experimental evaluator?

One final example of a prima facie problem. All of the normative theories have failed to mention the role of the structures of the subject matter in conducting evaluations. To what extent does learning mathematics (or morality) depend on the structure of what has to be learned? Does such a structure have to be unraveled by those who understand mathematics (or morality), with, perhaps, the help of philosophers?[38] What position do (and should) the various theories of evaluation assume concerning what counts as an explanation of human behavior?

I am suggesting, then, that an adequate normative theory of educational evaluation must take a justifiable stand with respect to these questions. I am not trying to settle such questions here, but I am claiming that such questions are important ones for an evaluation theorist to answer. To a large degree, such questions have been ignored. The proper resolution of such questions must be based on the foundations of evaluation theory--on the empirical and rational-ethical grounds. In the end, whether a given theory of education evaluation can be justified depends upon whether it satisfies the empirical tests and the rational-ethical tests. Failure to understand the proper role of the foundations can only lead to confusion, error, and wrongheadedness, in theory and in practice.

Implicit throughout the discussion has been the following proposition. The theories of evaluation which have been put forward by various writers need to be defended; the theoretical views need to be justified. It is wrongheaded merely to compare one theory of evaluation to another theory of evaluation. What is logically required are the grounds for determining the relative strengths and limitations of the theories. The strengths and limitations of a theory can only be determined by appeal to the relevant principles and the relevant kinds of reasons. The foundations constitute the proper grounds. All reasonable defenses of the various theories of evaluation must appeal ultimately to the foundations.

Cronbach and Ross have stated that

> As it becomes clearer what epistemological stances, what views of the political system, or what assumptions about the purposes of evaluation leads to each divergent position taken by evaluators, researchers, and decision-makers, babble may be replaced by the kind of argument that brings the field to maturity--maturity in service, as well as of internal clarity.[39]

As my previous remarks show, I have a general sympathy with their statement. But I would criticize them for their omission or lack of clarity on certain important questions. In particular, their failure to mention the rational and ethical questions is a serious oversight. Apparently it cannot be stated too often that educational evaluation is a practical activity which must be grounded on empirical tests and rational and ethical principles.

VI. Summary

The scope and depth of issues and problems that fall under the rubric of educational evaluation is enormous. It should be clear to all that some organization, systematization, and justification of the various views and statements about educational evaluation is urgently needed. I have tried to illustrate the very important way in which the claims of the field must be organized, systematized, and justified. In particular, I have tried to clarify and explicate the logical structure and the logical foundations of educational evaluation theories. I have shown that educational evaluation theories belong either to the descriptive (or empirical) realm or to the normative realm. I have shown that there are various forms of normative, educational evaluation activity, each of which has its distinctive goals, purposes, principles, standards of excellence, obligations, responsibilities, and duties. I have argued that such principles, criteria, and standards, and so on, need to be codified, organized, and systematized for each of the various forms of evaluation. I have argued that all theories must be founded on certain grounds. The foundations of educational evaluation theory are constituted

by empirical grounds and by rational-ethical grounds. Ultimately, the justification of a particular theory of evaluation depends on whether is satisfies the empirical and the rational-ethical tests.

A proper understanding of the logical structure of theories of educational evaluation, and especially an understanding of the logical foundations, will enhance our argumentation and discussion concerning educational evaluation. Such an understanding will eventually lead the field toward greater comprehensiveness, productivity, and defensibility in its theorizing and in its theory-guided practice. Failure to achieve such an understanding can only lead to confusion, error, and wrongheadedness.

MS15:d

Footnotes

[1]For an excellent discussion of the role of definitions in education, see Israel Scheffler, Chapter 1, The Language of Education (Springfield, Ill.: Charles C. Thomas, 1960). Scheffler's remarks about programmatic definitions are especially appropriate here. See also P. T. Gack, Chapter 9, Reason and Argument (University of California Press, 1976).

[2]Robert A. Dahl, Modern Political Analysis (2nd Ed.) (Englewood Cliffs, N.J.: Prentice-Hall, 1970), pp. 1-2.

[3]Paul A. Samuleson, Economics: An Introductory Analysis (5th Ed.) (New York: McGraw-Hill, 1961), p. 8.

[4]Ibid., p. 6.

[5]Richard Rudner, Philosophy of Social Science (Englewood Cliffs, N.J.: Prentice-Hall, 1966), p. 10.

[6]For example, see William L. Speizman, "Evaluation: An Evaluation from a Sociological Perspective. In The Uses of the Sociology of Education, NSSE Yearbook, 1974, Chapter 9, 192-210. Also see Carol H. Weiss, "Evaluation Research in the Political Context," in E. L. Struening and M. Guttentag (Eds.), Handbook of Evaluation Research (Vol. I) (Beverly Hills, Calif.: Sage Publications, 1974), pp. 13-26.

[7]Blaine R. Worthen and James R. Sanders, Educational Evaluation: Theory and Practice (Worthington, Ohio: Charles A. Jones, 1973), p. 1.

[8]Carol H. Weiss, p. 19. Are they asking whether evaluation is necessarily biased toward the status quo?

[9]Robert E. Stake, "The Countenance of Educational Evaluation." In Worthen and Sanders, 106-125, p. 111.

[10]For example, C. Weiss, p. 25.

[11]W. James Popham, Education Evaluation (Englewood Cliffs, N.J.: Prentice-Hall, 1975), p. 16. By the way, what counts as improving education?

[12]Worthen and Sanders, P. 349.

[13]Thomas D. Cook and Charles L. Gruder, "Metaevaluation Research," Evaluation Quarterly, February 1978, 2(1), 5-52.

[14]C. Weiss, P. 25.

[15]David K. Cohen, "Politics and Research: Evaluation of Social Action Programs in Education" Review of Educational Research, April 1970, 40 (2), 213-238, p. 236.

[16]Michael Quinn Patton, Utilization-Focused Evaluation (Beverly Hills, California: Sage Publications, 1978), p. 40.

[17]Ibid, p. 49.

[18]Ibid, P. 231.

[19]For example, Richard Brandt, Ethical Theory: The Problems of Normative and Critical Ethics (Englewood Cliffs, N.J.: Prentice-Hall, 1959), p. 5. Also William K. Frankina, Ethics (2nd Ed.) (Englewood Cliffs, N.J.: Prentice-Hall, 1973).

[20]For example, see David Cohen, "Politics and Research: Evaluation of Social Action Programs in Education, p. 233; Lee Ross and Lee J. Cronbach, editors of a review by A Task Force of the Standard Evaluation Consortium of the Handbook of Evaluation Research, edited by M. Guttentag and E. L. Struering, Proceedings of the National Academy of Education, (Vol. 3), 1976, pp. 81-107; Carol H. Weiss, "The Politization of Evaluation Research," in Cases in Public Policy-Making, edited by James E. Anderson (New York: Praeger Publishers, 1976), pp. 310-322, especially pp. 313-323.

[21]Thomas D. Cook and Charles L. Gruder, "Metaevaluation Research," Evaluation Quarterly, February 1978, 2(1), 5-52, p. 47.

[22]Scarvia B. Anderson, Samuel Ball, Richard T. Murphy, and Associates (Eds.), Encyclopedia of Educational Evaluation (San Francisco: Jossey-Bass Publishers, 1976), p. 348.

[23]Lee Ross and L. J. Cronbach (Eds.), "Book Review of Handbook of Evaluation Research," Proceedings of the National Academy of Education, 1976, 3, 81-107, pp. 100-102.

[24]W. James Popham, Educational Evaluation (Englewood Cliffs, N.J.: Prentice-Hall, 1975). See also, Norman E. Gronlund, Measurement and Evaluation in Teaching (2nd ed) (New York: MacMillan, 1971), and William A. Mehrens and Irvin J. Lehman, Measurement and Evaluation in Educational Psychology (New York: Holt, Rinehart and Winston, 1973).

[25]Louise L. Tyler and M. Frances Klein, et al., Evaluating and Choosing Curriculum and Instructional Materials (Los Angeles: Educational Research Associates, 1976), p. 21.

[26]Ibid, pp. 21-22.

[27]Ibid., p. 7.

[28]For further description of such evaluations, see Marvin Alkin, Richard Daillek, and Peter White, Evaluation Utilization in Local Schools: A Case Study Approach (Sage Publications, in press). It should be obvious that this kind of an evaluation would require instruction which goes beyond instruction in traditional research methodology. Certain theorists are trying to develop an instructional program appropriate for this form of evaluation. See also Michael Patton, Utilization-Focused Evaluation.

[29]Ross and Cronbach (1976), pp. 102-103.

[30]For example, see Marvin C. Alkin, "Evaluation Theory Development." Evaluation and Comment, 1969, 2 (1), 2-7.

[31]Anderson, Ball, Murphy, and Associates (1973), pp. 136-137. Robert E. Stake, "The Countenance of Educational Evaluation," in Worthen and Sanders, 106-125. Daniel L. Stufflebeam, Ego G. Guba, et al., Educational Evaluation and Decision-Making in Education (Ithaca, Ill.: Peacock, 1971).

[32]For a discussion of the logical difference between rationality and morality, see Kurt Baier, The Moral Point of View (New York: The Random House, 1965), pp. 93-96; R. M. Hare, Freedom and Reason (Oxford University Oress, 1963), 86 ff.; David A. J. Richards, A Theory of Reasons for Action (Oxford University Press, 1971), pp. 77-78 (Chapter 3, "The Principles of Rational Choice, and Chapter 6, "The Concept of Morality").

[33]Michael Scriven, "The Methodology of Evaluation" In R. E. Stake (Ed.), Curriculum Evaluation. American Educational Research Association Monograph Series on Evaluation, No. 1 (Chicago: Rand McNally, 1967), pp. 52-54.

[34]Michael Scriven, "The Concept of Evaluation," in M. W. Apple, M. J. Subkoviak, and H. S. Lufler, Jr. (Eds.), Educational Evaluations: Analysis and Responsibility, pp. 55-82. See pages 56, 63, 66. Scriven holds the mistaken view that "value judgments are simply a species of scientific judgment" (p. 56). See also "The Methodology of Evaluation," P. 81.

[35]Worthen and Sanders (1973), pp. 21, 25, 26. See also Paul Samuelson, Economics: An Introductory Analysis (5th ed.) (New York: McGraw-Hill, 1961), pp. 9-10.

[36]Paul Samuelson, Economics: An Introductory Analysis, pp. 9-10.

[37]Robert A. Dahl, Modern Political Analysis (2nd Ed.) (Englewood Cliffs, N.J.: Prentice-Hall, 1970), pp. 1, 100, 104.

[38]Richard S. Peters, "Chairman's Remarks" in "Part One: Psychology as the Science of Human Behavior." In S. C. Brown (Ed.), Philosophy of Psychology. New York: Barnes and Noble, 1974, p. 59.

[39]Ross and Cronbach (1976), p. 106.

MS15:d